

Οικονομικό Πανεπιστήμιο Αθηνών
Μάθημα: “Στατιστική”
Συνάντηση: 2η



Δρ. Τρύφωνας Λεμοντζόγλου

Τι κρατάμε από την πρώτη μας συνάντηση:

Ποιος είναι ο χαρακτήρας των δεδομένων μας
ως προς:

α) τη φύση τους (**ποσοτικά-ποιοτικά**)

β) την πηγή άντλησής τους
(**πρωτογενή-δευτερογενή**)

γ) την κίνησή τους στο χώρο (i) και το χρόνο (t)
(**διαστρωματικά, χρονολογικά και δεδομένα
πάνελ**)

Τι κρατάμε από την πρώτη μας συνάντηση:

Απόπειρα εξαγωγής των βασικών περιγραφικών στατιστικών (*descriptive statistics*):

- 1) μέση τιμή (ή μέσος όρος) *mean*
- 2) διάμεσος *median*
- 3) επικρατούσα τιμή *mode*
- 4) διακύμανση (ή διασπορά) *variance*

ΠΙΝΑΚΕΣ ΠΕΡΙΓΡΑΦΙΚΩΝ ΣΤΑΤΙΣΤΙΚΩΝ (ΠΑΡΑΔΕΙΓΜΑΤΑ ΜΕ ΠΡΑΓΜΑΤΙΚΑ ΔΕΔΟΜΕΝΑ)

1870

1879

MeanMax.Min.St. Dev.Obs.

MeanMax.Min.St. Dev.Obs.

1. Literacy Development

Total Literacy (%)	14.42%	46.49%	1.99%	0.08	350	16.32%	50.50%	0.71%	0.08	350
Male Literacy (%)	25.04%	77.68%	3.27%	0.11	350	27.78%	66.09%	1.37%	0.12	350
Female Literacy (%)	3.61%	42.53%	0.00%	0.06	350	4.21%	39.40%	0.00%	0.06	350
Gender Gap in Literacy	38.67	367.00	0.61	59.14	350	36.75	635.50	0.74	61.41	350

2. Student Attainment

Total Student Attainment (%)	4.63%	17.93%	0.41%	0.02	350	4.45%	24.34%	0.03%	0.02	350
Male Student Attainment (%)	8.18%	23.52%	0.69%	0.04	350	7.79%	31.95%	0.06%	0.04	350
Female Student Attainment (%)	1.08%	13.60%	0.00%	0.01	350	0.97%	18.92%	0.00%	0.01	350

3. Landownership Patterns

Access to Land (%)	14.53%	95.48%	0.00%	0.17	350	14.92%	93.16%	0.00%	0.18	350
--------------------	--------	--------	-------	------	-----	--------	--------	-------	------	-----

4. Delayed Industrialization

Stuck in Agriculture	16.44	219.50	0.01	25.53	350	12.92	317.50	0.02	29.01	350
----------------------	-------	--------	------	-------	-----	-------	--------	------	-------	-----

5. Marriage Patterns

Married-Single ratio (total population)	0.58	0.92	0.25	0.07	350	0.58	1.30	0.37	0.09	350
Married-Single ratio (male population)	0.53	0.88	0.16	0.07	350	0.53	1.28	0.21	0.09	350
Married-Single ratio (female population)	0.66	2.02	0.43	0.11	350	0.66	1.16	0.33	0.09	350

6. Women's Social Position

Female Student Participation	0.11	0.78	0.00	0.16	350	0.09	0.83	0.00	0.15	350
Female Teacher Participation	0.16	7.00	0.00	0.45	350	0.14	12.00	0.00	0.72	350
Female Labour Participation	0.89	90.05	0.00	6.84	350	0.25	16.5	0.00	1.38	350

	1907					1920				
	mean	max	min	std.dev.	obs.	mean	max	min	std.dev.	obs.
Female Employment	4.779%	12.232%	1.071%	0.028	66	11.047%	28.256%	3.494%	0.052	66
Modern Occupations	8.202%	24.614%	1.869%	0.050	66	12.954%	31.574%	4.789%	0.051	66
Manufacturing	10.902%	39.972%	1.616%	0.078	66	17.002%	44.493%	4.084%	0.078	66
Textiles	68.770%	100.000%	0.000%	0.326	66	68.343%	92.307%	35.678%	0.155	66
Self-Employment	16.954%	34.146%	3.141%	0.059	66	18.091%	42.105%	1.750%	0.084	66
Commerce/Trade	1.272%	15.488%	0.000%	0.026	66	2.418%	6.557%	0.343%	0.012	66
Municipal/Public Employees	0.642%	4.545%	0.000%	0.009	66	5.914%	21.084%	0.000%	0.047	66
Agricultural Activities	2.207%	10.133%	0.000%	0.020	66	9.750%	32.022%	0.823%	0.066	66
Livestock Production	4.537%	38.888%	0.000%	0.077	66	12.281%	62.264%	0.000%	0.124	66
Agriculture	1.983%	10.127%	0.000%	0.021	66	9.857%	31.681%	0.850%	0.065	66
Fishing/Hunting	0.172%	3.482%	0.000%	0.006	66	0.810%	10.000%	0.000%	0.018	66
Females in Religious Org.	8.517%	90.804%	0.000%	0.140	66	12.918%	69.406%	0.000%	0.148	66
Female Servants	55.368%	90.476%	9.183%	0.198	66	68.770%	97.500%	27.848%	0.153	66

Διακύμανση (δειγματική)

$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + (X_3 - \bar{X})^2 + \dots + (X_n - \bar{X})^2}{n-1} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Τυπική Απόκλιση (δειγματική)

$$s = \sqrt{s^2}$$

Τώρα που γνωρίσαμε τη διακύμανση ας σκεφτούμε το
εξής: σε ποια κοινωνία θα προτιμούσαμε να ζούμε;

Κοινωνία Α: μέση τιμή = 90.000 ευρώ
διακύμανση = 1,35

Κοινωνία Β: μέση τιμή = 40.000 ευρώ
διακύμανση = 0,001

κοινωνία Α			κοινωνία Β		
άτομα (i)	income (x)		άτομα (i)	income (x)	
1	500		1	500	
2	459		2	2580	
3	578		3	5680	
4	420		4	7920	
5	389		5	350	
6	601		6	201	
7	594		7	10350	
8	498		8	15630	
9	605		9	450	
10	427		10	420	
mean	507,1		mean	4408,1	

Με μοναδικό κριτήριο
το **μέσο μηνιαίο
εισόδημα** θα
προτιμούσαμε μάλλον
να ζούμε στην
κοινωνία Β (έχει **μέσο
εισόδημα ΠΟΛΥ
μεγαλύτερο**)

κοινωνία Α			κοινωνία Β		
άτομα (i)	income (x)		άτομα (i)	income (x)	
1	500		1	500	
2	459		2	2580	
3	578		3	5680	
4	420		4	7920	
5	389		5	350	
6	601		6	201	
7	594		7	10350	
8	498		8	15630	
9	605		9	450	
10	427		10	420	
mean	507,1		mean	4408,1	
std. dev.	82,58120583		std. dev.	5370,112195	

Με κριτήριο την τιμή
της **ΤΥΠΙΚΗΣ**
ΑΠΟΚΛΙΣΗΣ ; ; ;

Υψηλότερη τυπ.
απόκλιση =
Υψηλότερη ανισότητα

μετοχή Α		μετοχή Β	
ημερονημία	τιμή κλεισίματος	ημερονημία	τιμή κλεισίματος
1/1/91	1,2	1/1/91	2,1
2/1/91	1,6	2/1/91	6,9
3/1/91	1,4	3/1/91	7
4/1/91	1,4	4/1/91	1,2
5/1/91	1,1	5/1/91	0,4
8/1/91	1,3	8/1/91	1,2
9/1/91	1,4	9/1/91	4
10/1/91	1,5	10/1/91	7,5
11/1/91	1,2	11/1/91	0,5
μέση τιμή	1,33	μέση τιμή	3,4
τυπική απόκλ.	0,17	τυπική απόκλ.	2,98

Με κριτήριο τη **μέση τιμή** η επένδυση στη μετοχή Β είναι πιο ελκυστική !!!

Με κριτήριο τον κίνδυνο ; ; ;

Απόπειρα εξαγωγής των βασικών περιγραφικών στατιστικών (συνέχεια ...)

2010	Αρδευτικές αγροτικές εκτάσεις (%)
Albania (i=1)	$X(1,1) = 17$
Bulgaria (i=2)	$X(1,2) = 2$
Denmark (i=3)	$X(1,3) = 12$
France (i=4)	$X(1,4) = 5$
Greece (i=5)	$X(1,5) = 18$
Israel (i=6)	$X(1,6) = 33$
Malta (i=7)	$X(1,7) = 34$
Pakistan (i=8)	$X(1,8) = 56$
Serbia (i=9)	$X(1,9) = 1$
Tunisia (i=10)	$X(1,10) = 4$

$$\text{μέση τιμή} = (17+2+12 + \dots + 56+1+4) / 10 = 18,2$$

$$\text{διάμεσος: } 1 \quad 2 \quad 4 \quad 5 \quad 12 \quad 17 \quad 18 \quad 33 \quad 34 \quad 56$$

$$\delta = (12+17)/2 = 14,5$$

Επικρατούσα Τιμή : ΔΕΝ υπάρχει !!!

$$s^2 =$$

$$[(17-18,2)^2 + (2-18,2)^2 + \dots + (4-18,2)^2] / 9 =$$

$$= 319,06$$

$$s = 17,87 \text{ (δειγματική τυπική απόκλιση)}$$

2010	Ρυθμός αύξησης πληθυσμού (%)
Albania (i=1)	$X(2,1) = -0,50$
Bulgaria (i=2)	$X(2,2) = -0,66$
Denmark (i=3)	$X(2,3) = +0,44$
France (i=4)	$X(2,4) = +0,49$
Greece (i=5)	$X(2,5) = +0,13$
Israel (i=6)	$X(2,6) = +1,83$
Malta (i=7)	$X(2,7) = +0,49$
Pakistan (i=8)	$X(2,8) = +2,47$
Serbia (i=9)	$X(2,9) = -0,40$
Tunisia (i=10)	$X(2,10) = +1,08$

μέση τιμή

$$= (-0,50 - 0,66 + 0,44 + \dots + 2,47 - 0,40 + 1,08) / 10$$

$$= 0,537$$

διάμεσος: -0,66 -0,50 -0,40 0,13 **0,44**

0,49 0,49 1,08 1,83 2,47

$$\delta = (0,44 + 0,49) / 2 = 0,465$$

Επικρατούσα Τιμή: 0,49

s^2

$$= (-0,50 - 0,537)^2 + (-0,66 - 0,537)^2 + \dots + (1,08 - 0,537)^2 / 9$$

$$= 1,03$$

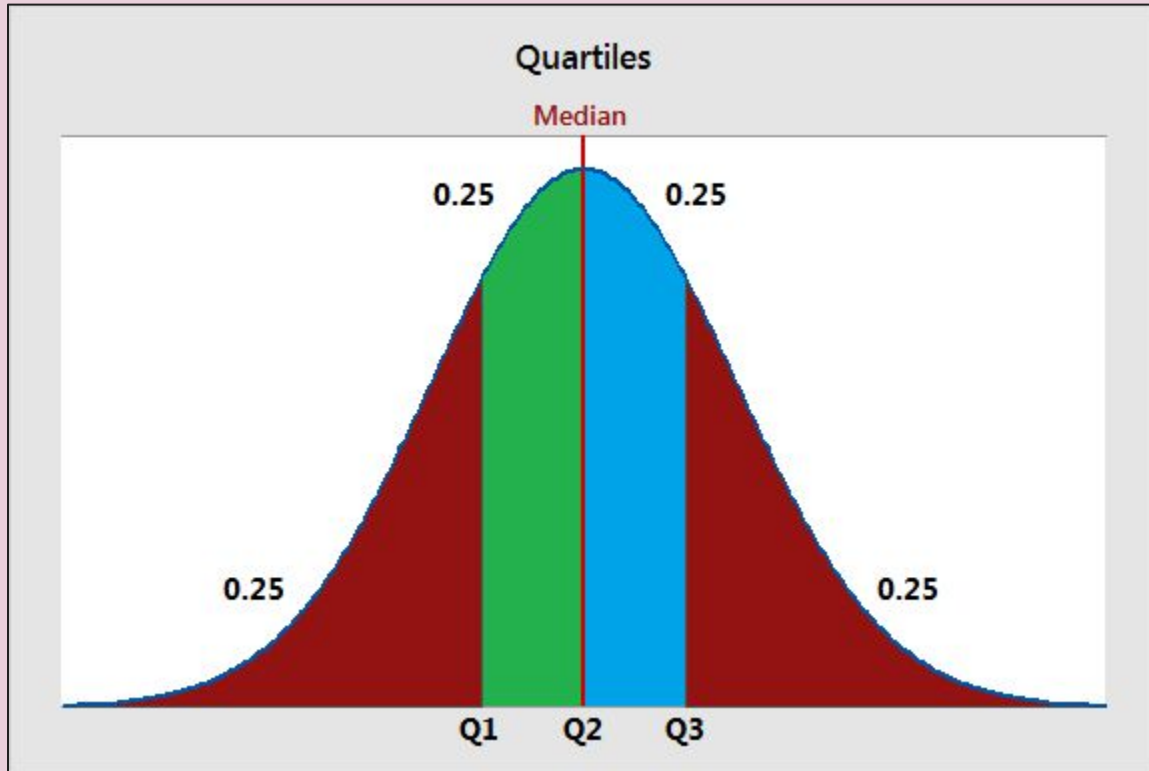
$s = 1,01$ (δειγματική τυπική απόκλιση)

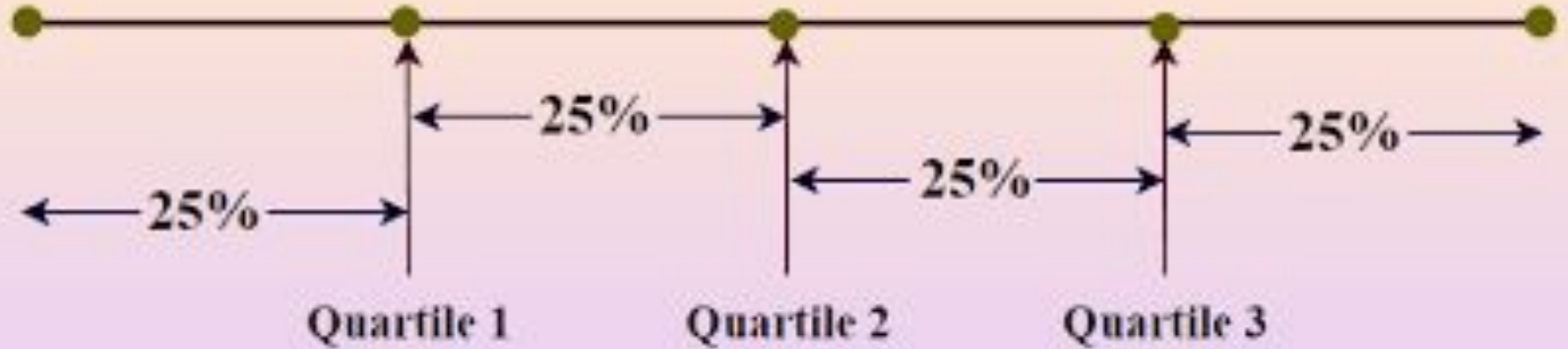
- **Τεταρτημόρια:** πρώτο, δεύτερο, τρίτο

Διάμεσος (ή δεύτερο τεταρτημόριο): αριστερά της υπάρχει το 50% των παρατηρήσεων του δείγματος και δεξιά της το υπόλοιπο 50% (χωρίζει το δείγμα μας σε δύο ίσα κομμάτια)

Πρώτο τεταρτημόριο: αριστερά του υπάρχει το 25% των παρατηρήσεων του δείγματος και δεξιά του το υπόλοιπο 75%

Τρίτο τεταρτημόριο: αριστερά του υπάρχει το 75% των παρατηρήσεων του δείγματος και δεξιά του το υπόλοιπο 25%





Q1 (πρώτο τεταρτημόριο):

είναι η τιμή που βρίσκεται στην θέση $(n + 1) / 4$

Q2 (δεύτερο τεταρτημόριο):

είναι η τιμή που αντιστοιχεί στην διάμεσο και βρίσκεται στη θέση $(n + 1) / 2$

Q3 (τρίτο τεταρτημόριο):

είναι η τιμή που βρίσκεται στην θέση $[3 * (n + 1)] / 4$

2010	Αρδευτικές αγροτικές εκτάσεις (%)
Albania (i=1)	$X(1,1) = 17$
Bulgaria (i=2)	$X(1,2) = 2$
Denmark (i=3)	$X(1,3) = 12$
France (i=4)	$X(1,4) = 5$
Greece (i=5)	$X(1,5) = 18$
Israel (i=6)	$X(1,6) = 33$
Malta (i=7)	$X(1,7) = 34$
Pakistan (i=8)	$X(1,8) = 56$
Serbia (i=9)	$X(1,9) = 1$
Tunisia (i=10)	$X(1,10) = 4$

Ας επιστρέψουμε στα αρχικά μας παραδείγματα ...

$$\text{μέση τιμή} = (17+2+12 + \dots + 56+1+4) / 10 = 18,2$$

διάμεσος: 1 2 4 5 12 17 18 33 34 56
 $\delta = (12+17)/2 = 14,5$

Επικρατούσα Τιμή : **ΔΕΝ** υπάρχει !!!

Καμία παρατήρηση δεν εμφανίζεται περισσότερες φορές συγκριτικά με τις άλλες.

$$\mathbf{Q1} = (10+1)/4 = 2,75 \rightarrow (3,5)$$

$$\mathbf{Q2} = (10+1)/2 = 5,5 \rightarrow (14,5)$$

$$\mathbf{Q3} = [3(10+1)]/4 = 8,25 \rightarrow (33,25)$$

2010	Ρυθμός αύξησης πληθυσμού (%)
Albania (i=1)	$X(2,1) = -0,50$
Bulgaria (i=2)	$X(2,2) = -0,66$
Denmark (i=3)	$X(2,3) = +0,44$
France (i=4)	$X(2,4) = +0,49$
Greece (i=5)	$X(2,5) = +0,13$
Israel (i=6)	$X(2,6) = +1,83$
Malta (i=7)	$X(2,7) = +0,49$
Pakistan (i=8)	$X(2,8) = +2,47$
Serbia (i=9)	$X(2,9) = -0,40$
Tunisia (i=10)	$X(2,10) = +1,08$

Ας επιστρέψουμε στα αρχικά μας παραδείγματα ...

μέση τιμή

$$= (-0,50 - 0,66 + 0,44 + \dots + 2,47 - 0,40 + 1,08) / 10$$

$$= 0,537$$

διάμεσος: -0,66 -0,50 -0,40 0,13 **0,44**
0,49 0,49 1,08 1,83 2,47
 $\delta = (0,44 + 0,49) / 2 = 0,465$

Επικρατούσα Τιμή: 0,49

Q1: θέση $(10+1)/4 = 2,75 \rightarrow (-0,4)$

Q2: θέση $(10+1)/2 = 5,5 \rightarrow (0,465)$

Q3: θέση $[3(10+1)]/4 = 8,25 \rightarrow (1,3)$

- **μέγιστη & ελάχιστη τιμή:** π.χ μέσο ετήσιο οικογενειακό εισόδημα (σε ευρώ)

κοινωνία Α: μέση τιμή 11.000 ευρώ, μέγιστη τιμή = 150.000 ευρώ, ελάχιστη τιμή = 5.000 ευρώ

κοινωνία Β: μέση τιμή 5.000 ευρώ, μέγιστη τιμή = 7.000 ευρώ, ελάχιστη τιμή = 4.000 ευρώ.

Που θα προτιμούσατε να ζείτε;

2010	Αρδευτικές αγροτικές εκτάσεις (%)
Albania (i=1)	$X(1,1) = 17$
Bulgaria (i=2)	$X(1,2) = 2$
Denmark (i=3)	$X(1,3) = 12$
France (i=4)	$X(1,4) = 5$
Greece (i=5)	$X(1,5) = 18$
Israel (i=6)	$X(1,6) = 33$
Malta (i=7)	$X(1,7) = 34$
Pakistan (i=8)	$X(1,8) = 56$
Serbia (i=9)	$X(1,9) = 1$
Tunisia (i=10)	$X(1,10) = 4$

Ας επιστρέψουμε στα αρχικά μας παραδείγματα ...

$$\text{μέση τιμή} = (17+2+12 + \dots + 56+1+4) / 10 = 18,2$$

$$\text{διάμεσος: } 1 \quad 2 \quad 4 \quad 5 \quad 12 \quad 17 \quad 18 \quad 33 \quad 34 \quad 56$$

$$\delta = (12+17)/2 = 14,5$$

Επικρατούσα Τιμή : **ΔΕΝ** υπάρχει !!!

Καμία παρατήρηση δεν εμφανίζεται περισσότερες φορές συγκριτικά με τις άλλες.

$$Q1 = (10+1)/4 = 2,75 \rightarrow (3,5)$$

$$Q2 = (10+1)/2 = 5,5 \rightarrow (14,5)$$

$$Q3 = [3(10+1)]/4 = 8,25 \rightarrow (33,25)$$

max: 56 min: 1

2010	Ρυθμός αύξησης πληθυσμού (%)
Albania (i=1)	$X(2,1) = -0,50$
Bulgaria (i=2)	$X(2,2) = -0,66$
Denmark (i=3)	$X(2,3) = +0,44$
France (i=4)	$X(2,4) = +0,49$
Greece (i=5)	$X(2,5) = +0,13$
Israel (i=6)	$X(2,6) = +1,83$
Malta (i=7)	$X(2,7) = +0,49$
Pakistan (i=8)	$X(2,8) = +2,47$
Serbia (i=9)	$X(2,9) = -0,40$
Tunisia (i=10)	$X(2,10) = +1,08$

Ας επιστρέψουμε στα αρχικά μας παραδείγματα ...

μέση τιμή

$$= (-0,50 - 0,66 + 0,44 + \dots + 2,47 - 0,40 + 1,08) / 10$$

$$= 0,537$$

διάμεσος: -0,66 -0,50 -0,40 0,13 **0,44**

0,49 0,49 1,08 1,83 2,47

$$\delta = (0,44 + 0,49) / 2 = 0,465$$

Επικρατούσα Τιμή: 0,49

Q1: θέση $(10+1)/4 = 2,75 \rightarrow (-0,4)$

Q2: θέση $(10+1)/2 = 5,5 \rightarrow (0,465)$

Q3: θέση $[3(10+1)]/4 = 8,25 \rightarrow (1,3)$

max: +2,47%

min: -0,66

- **Εύρος τιμών** = (μέγιστη τιμή – ελάχιστη τιμή)

κοινωνία Α: εύρος τιμών = $(150.000 - 5.000) = \dots$

κοινωνία Β: εύρος τιμών = $(7.000 - 4.000) = \dots$

Η απόσταση που χωρίζει το μέγιστο από το ελάχιστο εισόδημα είναι πολύ μεγαλύτερη στην κοινωνία Α

ενδοτεταρτημοριακό εύρος = $(Q3 - Q1)$

συντελεστής μεταβλητότητας (%)
= $(\text{τυπική απόκλιση} / \text{μέση τιμή}) * 100$

$<10\%$ → ομοιογενές δείγμα

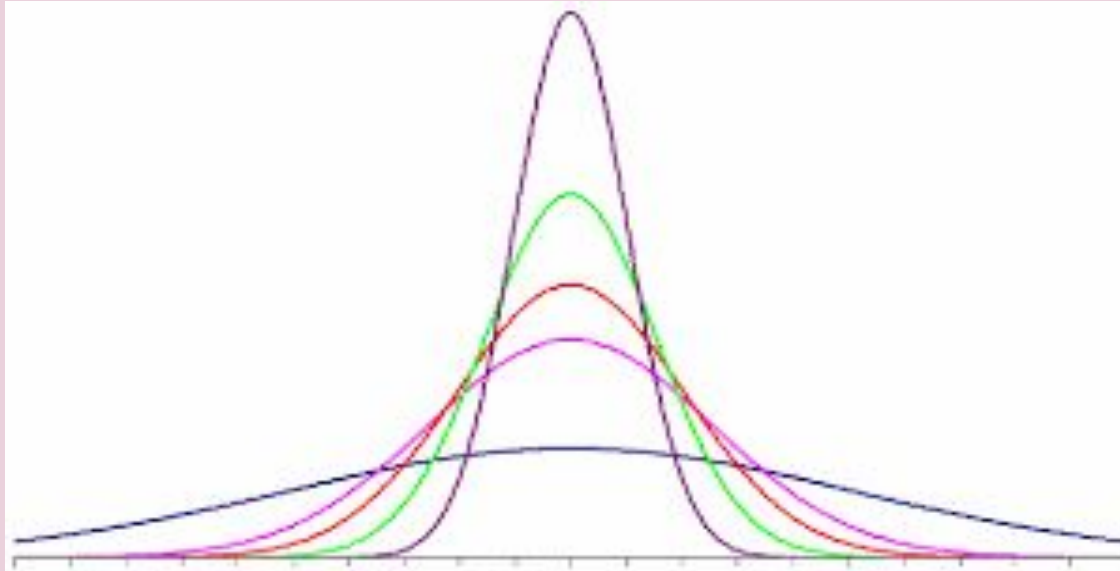
Συντελεστής ασυμμετρίας:

$$G = \frac{1}{(n-1)} \sum_{i=1}^k \left(\frac{x_i - \bar{x}}{s} \right)^3$$



Συντελεστής κύρτωσης:

$$K = \frac{1}{(n-1)s^4} \sum_{i=1}^n (x_i - \bar{x})^4$$



=3 (μεσόκυρτη)

>3 (λεπτόκυρτη)

<3 (πλατύκυρτη)